コンピュータと教育 68－6
（２００３．２．７）

日本語学習者の作文における
格助詞の誤り検出と訂正

今枝 恒治[†]　河合 敦夫[†]　石川 裕司[†]　永田 亮[†]　桝井 文人[†]

本論文では，実際に日本語学習者の書いた作文を題材として，比較的書き誤る可能性の高い格助詞に対して，誤りの検出と訂正を行うシステムを構築した．この手法として，まずルールに基づいた処理をおこない，そこで検出・訂正できなかった誤りに対して，格フレーム照合をおこなう．格フレーム照合では，入力文の格フレームと辞書から得られる格フレームを比較することにより，誤りの検出および訂正をする．実験の結果として，75.6%の誤り検出率と，62.5%の誤り訂正率を得ることができ，本手法が，日本語学習者の書いた作文中の格助詞の誤り検出・訂正に有効であることを示すことができた．

Error detection and correction of case particles in Japanese Learner's composition

KOJI IMAEDA[†], ATSUO KAWAI[†], YUJI ISHIKAWA[†], RYO NAGATA[†] and FUMITO MASUI[†]

In this paper, we propose a method of detecting and correcting errors in usage of case particles in composition written by Japanese learners. The method has two major steps.In the first step, rules based on analysis of the composition detect and correct the errors. In the second step, a case frame dictionary is used to detect and correct the errors that the rules failed to detect. Those errors are corrected comparing the case frame of an input sentence with the case frames in the dictionary. As a result of experiments, 75.6% of rate of error detection and 62.5% of rate of error correction were obtained. The result shows that the method is effective to detect and correct errors of case particles in composition written by Japanese learners.

[†] 三重大学工学部情報工学科
Department of Information Engineering,
Faculty of Engineering, Mie University

１．はじめに

近年，インターネットの普及に伴い，国際的な情報交換が日常のものとなり，世界的に，母語以外の言語，すなわち第二外国語を，コミュニケーションのために習得する必要と要望が急速に増してきた．また，日本語を読んで理解したり，書いたりする機会が増え，第二外国語として，日本語を学ぼうとする人々の数が，急激に増加している．

その需要に応えるためには，日本語教師の養成が必要であるが，人材の育成は短期間では難しい．特に，海外においては，日本語が母語である日本語教師だけでは，その必要は満たせない．そこで，日本語を母語としない人々が日本語を学び・教えられる環境を様々な形で支援する必要がある．

世の中では，計算機で第二外国語教育を支援するシステムとして，オーディオやビデオによる LL（Language Laboratory）が，まず開発された．その後，コンピュータ支援教育 CAI（Computer Assisted Instruction）の中でも特に，語学教育に特化した分野であり，学習者の自律的学習が注目され始めた．すなわちコンピュータやコンピュータネットワークを学習過程全般にわたり，学習者が主体的に活用していくということであり，その意味を持つコンピュータ支援言語学習 CALL（Computer Assisted Language Learning）という語が最近では使われている．CALL は，教室での授業と比べて，自分のペース・能力に合わせて学習ができるという特徴を持っている．

現状の日本語作文の授業では，日本語学習者が作文を書き，教師が添削することを繰り返すのが通例となっている．しかし，教師にとって，日本語学習者の作文を，一つ一つ添削することは，時間的な負担となる．そのため，教師の添削を支援したり，自動的に添削を行える CALL システムを構築することで，教師の時間的な負担，また，場所的な負担も軽減することができる．

以上より，本論文では，日本語学習者を対象として，書き誤る可能性の高い格助詞についての誤りを検出し，訂正するシステムを提案し，日本語学習者の学習能力の向上を目指した．

以下，２章で誤りの分類と関連研究を紹介し，３章で，システムの概要と構成について述べる．それに基づき，４章で実験を行い，その評価を行う．その時に得られた課題を５章で紹介し，６章で本稿についてまとめる．

２．誤りの分類と関連研究

２．１　誤りの分類

まず，日本語学習者が作文を書く上で，犯しやすい誤りを，文法書[1]を元に，分類したものを例と共に以下に示す．

・格助詞の誤り
　例：会議　を　参加しました．
・授受・使役・受身誤り
　例：先生は，私に説明して　あげました．
・数の単位表現に関する誤り
　例：ナイフが三　冊　あります．
・用言の活用・時制誤り
　例：私は，昨日温泉に　行く．
・呼応の不一致による誤り
　例：10 円しか，財布の中に　ある．
　　×

本論文では，日本語学習者特有の誤りであり，文中で必須格となりやすい格助詞の誤りを中心として，取り上げた．助詞は，文の構成上，必ず使用され，重要な役割を果たしていて，誤用によって，文の意味までも変化する品詞である．また意味・用法も多岐に渡るので，日本語学習者にとっては習得し難い分野であるといえる．

２．２　関連研究

日本語学習者の作文の誤りを題材に，誤りを検出し，訂正する研究は既に，いくつか存在する．

掛川ら[2]は，日本語学習者の作文に対して，正誤だけではなく，学習者がどのように誤っていて，どう直すべきかを教えることを目的としたシステムを提案している．そこでは，学習者は具体的な

場面設定・係り関係に関する情報・使用できる語が限定されているなどの制約のもとで作文を行っている．この問題点として，自由作文に対応していないことが挙げられる．

橋本ら[3]は，助詞の誤りに注目して，その誤りの判断基準を動詞としている．用いるツールとして，IPAL の動詞辞書[4]と名詞辞書[5]を用いている．IPAL の辞書は，登録語数が，重要基本単語のみであり，少ない．このため，辞書に記載されていない単語が出てきた場合には使えない．また，対象としている助詞が，動詞の直前の助詞であることから，それ以外の助詞の誤りについては，対応しきれない．

納富ら[6]は，あらかじめ決められた構文パターンの例文を作り，それに，非形態素となるランダムな文字列を含ませて，誤文を作成し，実験を行った．また実験結果では，訂正候補文までは，提示しているが，絞込みまでは，言及されていない．

本論文では，実際の日本語学習者の書いた作文を収集[1][7]し，NTT 日本語語彙大系[9][10]を用いて，文中のどの位置にある格助詞に対しても，もし誤りが存在すれば，検出し，訂正するシステムを提案する．本手法で用いる NTT 日本語語彙大系の特徴は，意味属性の種類が多く体系化されていることと，辞書に登録されている単語数，構文パターンが，他の辞書よりも多く，利用しやすい点にある．

3．システムの概要と構成

3．1　システムの概要

本システムでは，日本語学習者が入力した文に対して，その文が正しければ，そのまま正文を表示し，誤っていれば，誤り箇所と，訂正結果を表示する．ただし，訂正候補文が複数ある場合には，それらを全て候補として出力する．

3．2　システムの構成

始めに，本システムの構成を図１に示す．
次に各処理について説明する．

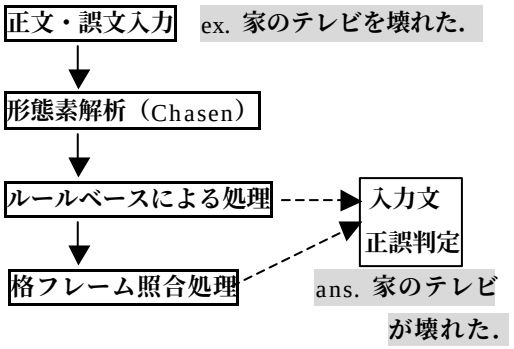

図１　システムの構成

①正文・誤文入力

入力文の種類としては，正文と誤文の両方を使用する．制約として，単文のみを扱い，誤りが存在する場合には，それは格助詞の部分で，一箇所だけであると仮定する．

②形態素解析

本研究では，形態素解析器として，茶筅[8]を用いて，品詞コードで出力させたものを利用する．

③ルールベースによる処理

②での出力結果と，品詞を利用したルールを用いることにより誤りを検出する．ルールベースを用いた処理の具体例を図２に示す．

この例では，名詞と用言（この例では動詞）が連続していることがわかる．そこで，ルールを適用し，その間にはさまれるべき助詞が脱落しているという誤りを検出する．
図２のルールの他に，

- 同じ助詞が連続している文
  例：兄が，大学[をを]卒業した．
- 助詞「を」が２つ含まれている文
  例：私は，テニス[を]練習[を]する．
- 隣接できない助詞を使用している文
  例：僕は，それ[しか] [を]食べない．

以上の誤りが，この処理で検出できる．

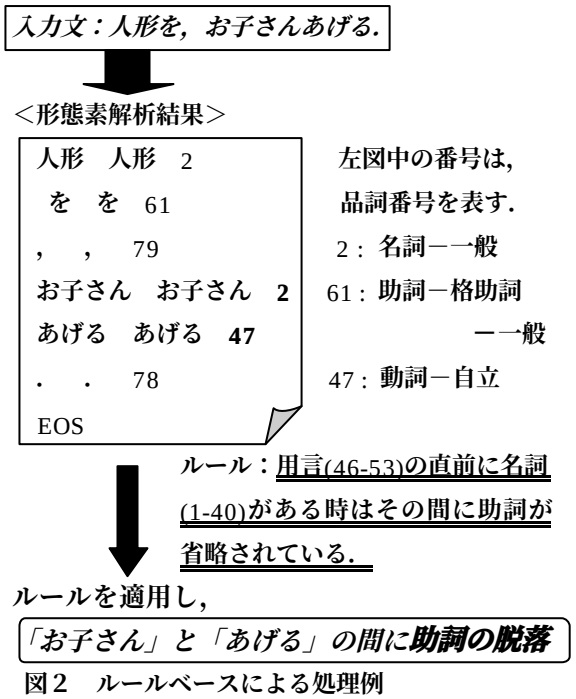

図２　ルールベースによる処理例

④格フレーム照合処理

この処理では，まず２つ（a，b）の処理を別々に行う．それらの結果をｃで，照合させることによって，入力文が正文であれば，正文を表示し，入力文が誤りを含むならば，その誤りを検出して，訂正まで行う処理である．

以下で，a．b．ｃについて説明する．

a．入力文中の名詞の意味属性の決定

意味属性とは，名詞の持つ意味を表すものである．まず，入力文から，含まれる名詞を全て抽出する．これらの名詞の意味属性の決定は，NTT 日本語語彙大系：Vol.１意味体系[9]との比較により行う．この具体例を図３に示す．

b．入力文中の用言を元に，格フレームを抽出

ここでは入力文中の用言と，NTT 日本語語彙大系：Vol.５構文体系[10]（以後，構文体系辞書と呼ぶ.）の比較を行う．構文体系辞書からは，用言の格フレーム，すなわち取り得る格（助詞）の情報と，その格の取り得る名詞の意味属性が得られる．図３の入力文を用いたこの処理を図４に示す．

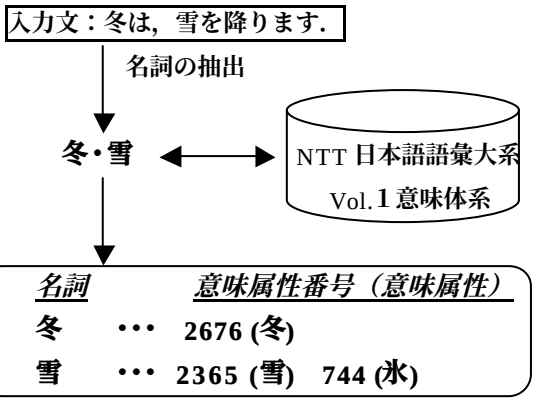

図３　意味属性の決定までの流れ

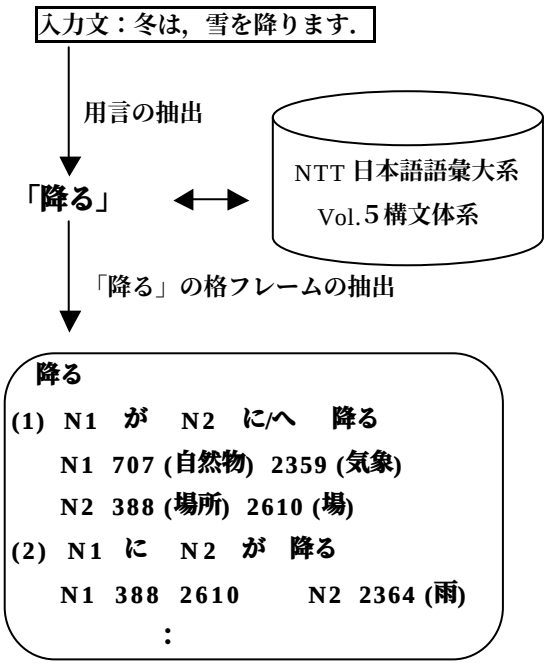

図４　用言から，格フレームを得るまでの流れ

c．ａとｂで得られた結果の照合

ａとｂの結果を元にした照合方法を図５に示す.

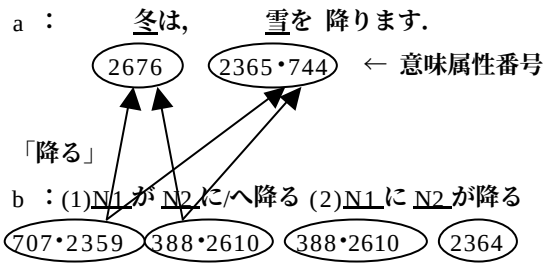

図５　意味属性と用言の格フレームとの照合方法

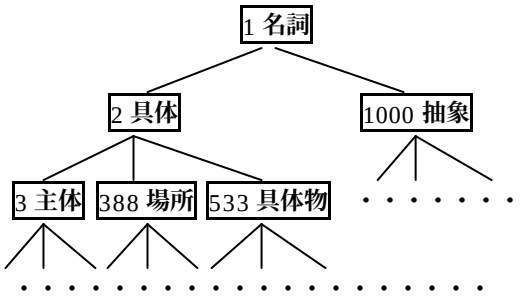

図６　名詞意味属性の木構造

（NTT 日本語語彙大系 Vol.１意味体系より）

図５において，b の構文体系辞書から得られた意味属性と，a の入力文中の名詞の意味属性を比較する．比較の際に，図６の名詞意味属性の木構造を用いる．この木構造は深さ 12 段から成り，約 2700 個の意味属性で構成されている．例えば，意味属性 3 は，3 から 387 までの意味属性を指す．

図５の例では，b の意味属性 707 と a の意味属性 2676 を最初に比較するが，意味属性 707 は 707 から 759 の意味属性は含むが，意味属性 2676 は含まない．次に意味属性 707 と意味属性 2365 を比較する．このようにして，すべての b の意味属性と a の意味属性を照合する．照合結果を図４で得られたパターン別に，表１に示す．

（1）

b＼a		冬	雪	
		2676	2365	744
N1	707	×	×	○
	2359	×	○	×
N2	388	×	×	×
	2610	×	×	×

（2）

b＼a		冬	雪	
		2676	2365	744
N1	388	×	×	×
	2610	×	×	×
N2	2364	×	×	×

表１　a と b の照合の結果

入力文　；　*冬は，雪を降ります.*

候補文

冬は，雪 が 降ります.　｝（１）より

冬は，雪 が 降ります.

候補文なし.　｝（２）より

図７　表１より得られた訂正候補文

この表１で，丸印は，入力文中の名詞の意味属性と，構文体系辞書での意味属性が一致したものを示している．丸印となったそれぞれの意味属性の後ろの助詞を比較する．比較した結果，助詞が同じ場合には何もせず，異なっていた時に，入力文中の助詞を構文体系辞書の助詞に変換する．

この例では，（１）で，N1 の 707 は入力文中の“雪”の意味属性 744 と一致することがわかる．従って，図４で得られた格フレームの N1 の後に続く助詞の“が”を，入力文中の“雪”の後に付加することによって，1 行目の訂正候補文を得る．２行目の訂正候補文は（１）の N1 の２つ目の意味属性 2359 と入力文中の“雪”の意味属性 2365 が一致することにより得られる．この流れで得た訂正候補文を，図７に示す．

４．実験

４．１　実験データ

実験データとして用いたのは，参考文献[1][7]の中の８０誤文である．また，3 章で述べたように，入力文は正文も扱うので，上で抽出した，８０誤文を人手により正文に直した．この８０正文も入力文として使用する．従って，今回，正文と誤文，合わせて１６０文を入力文として用いた．

４．２　実験条件

入力文は，４．１で述べた１６０文を用いる．ただし，この入力文は，受身・使役・敬語表現は含まない文で，重文・複文を除外した単文である．また，辞書の登録語は，今回の実験では，入力文

＜誤入力文＞

	⓪	①	②	③	④	⑤	⑥	⑦	⑧
1位	In(=Cor)	In	In	Cor	×	×	Cor	×	×
2位以降		Cor	×	In	In Cor	In	×	Cor	×
該当数	0	3	7	11	8	4	26	17	4

In･･･Input　　Cor･･･Correct

×･･･Input でも Correct でもない候補文

＜正入力文＞

	⑨	⑩	⑪
1位	In (=Cor)	×	×
2位以降	×	In (=Cor)	×
該当数	35	39	6

In･･･Input　　Cor･･･Correct

×･･･Input でも Correct でもない候補文

表２　訂正候補文の出力からみた実験結果

に出てくる単語のみとした．名詞の意味属性は，意味体系辞書を参考に約１５０個，動詞の数は，構文体系辞書を参考に６２語を用意した．

４．３　実験結果

実験により得られた結果を，訂正候補文の出力の分類と共に表２に示す．この表２では，誤入力文８０文で実験した結果と，正入力文８０文で実験した結果を示した．表２の Input は，入力文を示していて，Correct は人間が正しいと判定できる文を示している．表２の各表で，Input と Correct が，それぞれシステムが出力する訂正候補文の中の 1 位に入っているのか，2 位以降に入っているかのどちらかで⓪～⑪に分類した．

1 位と 2 位以降で分けた根拠は，４．４節の後半で述べる．

前章の図７の例は，入力文（In）はシステムの出した訂正候補文中に無く，人間が正しいと判定できる文（Cor）が１番目に出てくることから，分類⑥に当てはまることがわかる．表の該当数とは，各分類に当てはまる入力文の数を示している．
この表で，誤入力文の場合の分類⓪を扱わない理由は，誤入力文（In）が正しい文（Cor）には，なり得ないからである．

４．４　評価方法

ここで，どのような場合に，入力文の誤りの検出・訂正が成功したかについて定義しておく．これについては，本研究が日本語学習者に対する誤り訂正支援であることを考慮に入れて，定義した．つまり，正しい文を入力した時に，誤っていると診断されたり，誤った文を入力した時に，正解であると診断されると，日本語学習者にとって，このシステムの意味がなくなってしまう．

まず，入力文の誤りの検出について定義する．誤入力文の誤りの検出は，入力文と同じ文が，訂正候補文の中になければ，入力文が，誤っていたとわかることから，表２の⑥・⑦・⑧が検出に成功したとする．また，正入力文の正解の検出の場合は，入力文と同じ文が，正解候補文の中に，含まれていれば，入力文が正しいとシステムの使用者が理解できることから，⑨・⑩が検出に成功したとする．

次に，訂正の場合を定義する．誤入力文での誤りの訂正は，入力文が訂正候補文の中に含まれずに，1 位に正しく直された文があるとき成功とした．表２では⑥がそれに該当する．⑦・⑧を排除することにより，判定を厳しくし，日本語学習者に対する誤り訂正としての本来の精度が得られるからである．また，正入力文の場合は，正解候補文の中に，入力文があれば，入力文はもともと正しかったと，システムの使用者が，理解できることから，検出の場合同様，表２の⑨・⑩がそれに該当する．

上の定義に従って，検出率・訂正率を求めると，

検出率･･･（４７＋７４）／１６０＝７５．６%

訂正率･･･（２６＋７４）／１６０＝６２．５%

という，結果が得られた．

５．考察

実験結果では，構文体系辞書の格フレームと照合したものをすべて表示させていて，訂正候補の絞込みまでは行っていない．そのため，実際に日本語学習者が使用する際に，正入力文の場合には，入力文と同じ文を見つければよい．しかし，誤入力文の場合は，表２の①・④・⑦のように，訂正候補の中に正解はあるけれども，それが２位以降に存在するので，この中のどの文が正解であるかわからない．このため，学習者は訂正候補の中から，正解を選択するという事態となってしまう．これを避けるためには，正解文を 1 位へ持ってくる方法を考える必要がある．

５．１　訂正候補文中に正解がある場合

構文体系辞書は，頻度順に載せられているので，出てくる順にコストを低く割り振るという案が考えられる．しかし，これだけでは，結果は変わらない．そこで，構文体系辞書で使用されている“＊”について考えてみる．構文体系辞書では，“＊”の定義を，すべての意味属性としているので，必ず格フレーム照合処理で一致してしまい，訂正候補文が増える原因となっている．そこで，“＊”のコストを低く設定することにより，頻度順位ではそれより下の構文パターンが，より上位に来るようにする方法である．また，別の方法として，複数訂正候補文が出力された場合に，同じ文がその中にいくつか含まれる場合があるので，訂正候補文の中から，多数決により一つに絞り込むという方法も考えられる．

５．２　訂正候補文中に正解がない場合

次に，表２の②・⑤・⑧・⑪のように，訂正候補の中に，正解を含まない入力文について検討する．これらは，難易度により，大きく３つに分けることができる．

(１) 改良により訂正可能な誤文

例：私を人間です．（⑧）　→　**だ文**

構文体系辞書には，「です」に関する格フレームがない．通常，使われる格フレームは，

（名詞）は／が（名詞／形容詞）です.

が，大半であるので，個別に登録すれば訂正できる．

(２) 辞書のみでは解析できず，ルールが存在すれば訂正可能な誤文

例：タバコ屋の前に会う．（②）

新しい仕事を，一ヶ月からやめた．（⑤）

→　**任意格**

この例は，構文体系辞書から格の情報が得られない任意格を含む誤文の例である．任意格とは，手段や理由など，文の構成上，付加的な格要素のことをいう．解析手段として，

期間や場所の後には，「で」がつく傾向が強い.

というルールを作り，適用すれば，今回用いた入力文ではできそうである．しかし，このルール自体が，文法書などに，記載されておらず，確証性は持てないので今後検証していく予定である．

次に別の例を示す．

例：ワープロが使用方法を教えた．（②）

→**「の」**

これは，助詞の前の名詞が助詞の後の名詞に係る場合の例である．この時，構成が，

一般名詞　＋　助詞　＋　サ変接続名詞

であるので，この構成の場合は，

助詞の位置に「の」を挿入する

というルールを作ればできそうだが，これも確証性がないので，今後検証していく予定である．

(3) 辞書のみでは,解析できない誤文

例：娘は，家を嵐を引き起こす．(⑧)

その本に，神を書く．(⑧)

彼女は，誰もにあいさつしなかった．(⑧)

→ 助詞の入れ替え

一番上の例は,「家を」を「家に」に変換する例であるが，後続する「嵐を引き起こす」が特殊な表現であり，構文体系辞書にも記載されておらず，解析が困難な例である．

真ん中の例は，本来,「を」を「と」に変換する例であるが,「神」が書いた内容であることがわからないと解析できない例である．もし「」（カギ括弧）のような情報があれば，できる確率は上がると考えられる．

一番下の例は，本来「も」と「に」が入れ替わったものであるが,「誰もに」という句は日本語で用いることができ，解析するには，詳細な意味解析が必要になると考えられる．

以上から，訂正候補の中に正解がある場合に，それを上位にする方法と，訂正候補中に正解を含んでいない場合に３つの難易度に分類し，解説したが，今後，更なる考察が必要である．

６．まとめ

本論文では，日本語学習者の書いた作文を題材にして，ルールベースによる処理と格フレームを用いた意味処理を用いて，格助詞の誤りを検出し，訂正する手法を提案した．実験の結果として，約７６％の検出率と約６３％の訂正率を得ることができ，本手法が格助詞の誤り検出と訂正に有効であることが示せた．

今後の課題として，

- 検出と訂正の比率を上げるための解析手法を提案し，システム化．
- 異なる入力文・異なる助詞を用いての実験．
- 実際に日本語学習者が運用できるようなインタフェースの構築．
- 訂正候補だけでなく，なぜ誤りであるかの情報の付与．

を挙げることができ，実現できれば，システムの有効性も上がる．

参考文献

[1] 明治書院企画編集部編，日本語誤用分析，明治書院

[2] 掛川淳一，神田久幸，藤岡英太郎，伊丹誠，伊藤紘二,“日本語学習支援システムにおける作文診断処理系の提案と試作”，電子情報通信学会論文誌 D-I Vol.J83-D-I No.6 2000 pp.693-701

[3] 橋本利典，島田静雄,“外国人の書いた文章の助詞使用誤りの抽出”，情報処理学会 NL 研 117-2 1997.1 pp.9-14

[4] 情報処理振興事業協会，計算機用日本語基本動詞辞書 IPAL（Basic Verbs），1987

[5] 情報処理振興事業協会，計算機用日本語基本名詞辞書 IPAL（Basic Nouns），1996

[6] 納富一宏，石井博章,“日本語文書における共起格情報を用いた助詞要素の訂正”，神奈川工科大学研究報告 B-23， 1999

[7] 市川保子,“日本語誤用例文小辞典”，凡人社

[8] 松本祐治，北内啓,“日本語形態素解析システム 「茶筅」 version2.0 使用説明書”

[9] 池原悟，宮崎正弘，白井諭，横尾昭男，中岩浩巳，小倉健太郎，大山芳史，林良彦，NTT 日本語語彙大系 Vol.1. 意味体系, 岩波書店 1997

[10] 池原悟，宮崎正弘，白井諭，横尾昭男，中岩浩巳，小倉健太郎，大山芳史，林良彦，NTT 日本語語彙大系 Vol.5．構文体系，岩波書店 1997